The Patent Office

PCT/GB 9 9 / 0 2 7 1 6

2/3



4

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

0 9 7 0 3 0 7 6

GB99/2716

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed P. Mahoney

Dated 7 September 1999

**PRIORITY DOCUMENT**

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Patents Form 1/77

Patents Act 1977
(Rule 16)

THE PATENT OFFICE
18 AUG 1998
RECEIVED BY HAND

The Patent Office

1/77

18 AUG 1998

THE PATENT OFFICE
J
18 AUG 1998
RECEIVED BY HAND

# Request for grant of a patent

*(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)*

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

PPD 50348/GB/P

19AU098 E384219-1 D02093
P01/7700 25.00 - 9818001.1

2. Patent application number
*(The Patent Office will fill in this part)*

**9818001.1**

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

ZENECA Limited
15 Stanhope Gate
London W1Y 6LN
UNITED KINGDOM

Patents ADP number *(if you know it)*

6254007002

If the applicant is a corporate body, give the country/state of its incorporation

UNITED KINGDOM

4. Title of the invention

GENETIC METHOD

5. Name of your agent *(if you have one)*

Frank Mackie HUSKISSON

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Intellectual Property Department
ZENECA Agrochemicals
Jealott's Hill Research Station
P O Box 3538
Bracknell Berkshire RG42 6YA
UNITED KINGDOM

Patents ADP number *(if you know it)*

03969060002

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

| Country | Priority application number *(if you know it)* | Date of filing *(day / month / year)* |
|---|---|---|

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

| Number of earlier application | Date of filing *(day / month / year)* |
|---|---|

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

*a) any applicant named in part 3 is not an inventor, or*

*b) there is an inventor who is not named as an applicant, or*

*c) any named applicant is a corporate body.*

*See note (d))*

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

| | |
|---|---|
| Continuation sheets of this form | |
| Description | 29 |
| Claim(s) | 4 |
| Abstract | — |
| Drawing(s) | 14 + 14 |

10. If you are also filing any of the following, state how many against each item.

| | |
|---|---|
| Priority documents | |
| Translations of priority documents | |
| Statement of inventorship and right to grant of a patent *(Patents Form 7/77)* | |
| Request for preliminary examination and search *(Patents Form 9/77)* | |
| Request for substantive examination *(Patents Form 10/77)* | |
| Any other documents *(please specify)* | |

11. I/We request the grant of a patent on the basis of this application.

ZENECA LIMITED

Signature W Kallipeddi Date 18th August 1998

Authorised Signatory

12. Name and daytime telephone number of person to contact in the United Kingdom

VIJAYA KUMARI MALLIPEDDI 01344 414365

**Warning**

*After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.*

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

# GENETIC METHOD

The present invention relates to a method for the coexpression of two or more proteins in plants within a single transcription unit, to linker sequences for use in the method of the invention, to DNA constructs for use in the invention and to plants transformed with the constructs of the invention.

For many applications based on genetic modification of plants by transgenesis, it is desirable to express co-ordinately two or more transgenes. For instance, coexpression in plants of transgenes encoding antimicrobial proteins with different biochemical targets can result in enhanced disease resistance levels, resistance against a broader range of pathogens, or resistance that is more difficult to overcome by mutational adaptation of pathogens. Other examples include those aimed at producing a particular metabolite in transgenic plants by coexpression of multiple transgenes that are involved in a biosynthetic pathway. There are different ways to obtain transgenic plants expressing multiple transgenes. One frequently chosen option is to introduce each transgene individually via separate transformation events and to cross the different single-transgene expressing lines. The drawback of this method is that the different transgenes in the resulting progeny will be inserted at different loci, which complicates the subsequent breeding process. Moreover, this method is not applicable to crops that are propagated vegetatively, such as for instance potato, many ornamentals and fruit tree species. A second possibility is to introduce the different transgenes as linked expression cassettes, each with their own promoters and terminators, within a single transformation vector. Such a set of transgenes will in this case segregate as a single genetic locus. It has been observed, however, that the presence of multiple copies of the same promoter within a transgenic plant often results in transcriptional silencing of the transgenes (Matzke, M.A. and Matzke, A.J.M., 1998, Cellular and Molecular Life Sciences 54, 94-103). In an attempt to introduce a vector containing four linked transgenes each driven by a CaMV35S promoter, Van den Elzen P.J. *et al.* (Phil. Trans. R. Soc. Lon. B., 1993, 342: 271-278) observed that none of the analysed transgenic lines expressed all four transgenes at a reasonably high level. To avoid this problem one could use different promoters for each of the expression cassettes used in the construct. However, there is currently only a very limited choice of promoter sets that have comparable characteristics in terms of expression levels, cell-type and developmental specificity and response to environmental factors. A third option would be to produce

multiple proteins from one transcription unit by separating the distinct coding regions by so-called internal ribosomal entry sites, which allow ribosomes to reiterate translation at internal positions within a mRNA species. Although internal ribosomal entry sites are well documented in animal systems (Kaminski A. *et al.*, 1994, Genet. Eng. 16, 115-155) it is not known at present whether such sites are also functional in nuclear-encoded genes from plants. Polycistronic genes can be expressed when inserted in plant chloroplastic genomes (Daniell H. *et al.*, 1998, Nature Biotechnology 16, 345-348) but the gene products in this case are confined to the chloroplast, which is not always the preferred site of deposition of foreign proteins. A fourth strategy, finally, is based on the production of multiple proteins by proteolytic cleavage of a single polyprotein precursor encoded by a single transcription unit. Potyviruses, for instance, translate their genomic RNA into a single polyprotein precursor that encompasses proteolytic domains able to cleave the polyprotein precursor *in cis* (Dougherty, W.G. and Carrington, J.C., 1988, Annu. Rev. Phytopathol. 26, 123-143). Beck von Bodman, S. *et al.*, (1995, Bio/Technology 13, 587-591) have already exploited the potyviral system to co-express two enzymes involved in the biosynthesis of mannopine. The two biosynthetic enzymes were fused within one open reading frame together with a protease derived from a potyviral polyprotein precursor, and the adjoining regions were separated by 8 amino acids long spacers representing specific cleavage sites for the protease. The plants transformed with this construct synthesized mannopine, suggesting that the two enzymes had somehow been produced in a form that was at least partially functional, although direct evidence for the presumed cleavage events *in planta* was not presented. A disadvantage of this system is that a viral protein needs to be co-expressed with proteins of interest, which is not always desirable. More recently, Urwin P.E. *et al.* (1998, Planta 204, 472-479) have shown that it is possible to co-express two different proteinase inhibitors joined by a protease-sensitive propeptide derived from a plant metallothionein-like protein. A polyprotein precursor consisting of a cysteine protease inhibitor (oryzacystatin from vice), a propeptide from pea metallothionein-like protein and a serine protease inhibitor (cowpea trypsin inhibitor), was found to be cleaved in transgenic *Arabidopsis thaliana* plants. The cleavage, however, was only partial, as uncleaved polyprotein precursor could also be detected in the transgenic plants. As the polyprotein precursor did not contain a leader peptide, the translation products are predicted to be deposited in the cytosol. The metallothionein from which the propeptide was derived also does not contain a leader peptide (Evans IM 1990, FEBS Lett. 262, 29-32) and hence its processing

must occur in the cytosol. For some applications, cytosolic processing and deposition is a drawback. Many proteins, especially glycosylated proteins or proteins with multiple disulfide bridges, must be synthesized in the secretory pathway (encompassing the endoplasmic reticulum and Golgi apparatus) in order to be folded in a functional form (Bednarek and Raikhel 1992, Plant Mol. Biol. 20, 133-150). In addition, for some applications such as for instance the expression of antimicrobial proteins, the extracellular space is the preferred deposition site, as most microorganisms occur at least during the early stages of infection in the extracellular space. Proteins destined to the extracellular space are also synthesised via the secretory pathway but lack additional targeting information other than the leader peptide (Bednarek and Raikhel 1992, Plant Mol. Biol. 20, 133-150).

The present invention provides a convenient and highly efficient method of co-expressing two or more proteins in a plant as a single transcription unit where the two proteins are joined by a cleavable linker, the construct being designed such that cleavage occurs in the secretory pathway of the plant thereby releasing the proteins extracellularly.

According to the present invention there is provided a method for the expression of multiple proteins in a transgenic plant comprising inserting into the genome of said plant a DNA sequence comprising a promoter region operably linked to a signal sequence said signal sequence being operably linked to two or more protein encoding regions and a 3'-terminator region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide said propeptide providing a cleavage site whereby the expressed polyprotein is post-translationally processed into the component protein molecules.

The two or more protein encoding regions according to all aspects of the invention preferably do not encode identical proteins i.e. the method of the invention allows the production of different proteins in a single transcription unit.

The method for the expression of multiple proteins described herein does not cover the use of a linker propeptide derived from the Ib-AMP gene separating three protein encoding regions each of which encodes Rs-AFP2 and the insertion thereof into a plant genome.

As used herein the term signal sequence is used to define a sequence encoding a leader peptide that allows a nascent polypeptide to enter the endoplasmic reticulum and is removed after this translocation.

The signal sequence may be derived from any suitable source and may for example be naturally associated with the promoter to which it is operably linked. We have found the use

of signal sequences from the class of plant proteins known as defensins (Broekaert et al, 1995 Plant Physiol 108, 1353-1358; Broekaert et al, 1997, Crit, Rev, Plant Sci. 16, 297-323) to be particularly suitable for use in the method of the invention.

The promoter sequence may for example be that naturally associated with the signal sequence, and/or it may be that naturally associated with the protein encoding sequence to which it is linked, or it may be any other promoter sequence conferring transcription in plants. It may be a constitutive promoter or it may be an inducible promoter.

The linker propeptide may be a peptide which naturally contains processing sites for proteases occuring in the secretory pathway of plants, such as the internal propeptides derived from the Ib-AMP gene which are described further herein, or may be a peptide to which such a protease processing site has been engineered at either or both ends thereof to facilitate cleavage of the sequence. Where a propeptide possesses one such protease processing site a further protease processing site may be added. For example, as described fully herein, a further protease processing site has been added to the 3' end of the DNA sequence coding for the C-terminal propeptides from Dahlia and Amaranthus which naturally possess a protease processing site at their N-terminal end for an unknown secretory pathway protease.

In the present invention, we have developed two novel strategies for making artificial polyprotein precursors which are cleaved in the secretory pathway. The first one was based on the use of a propeptide derived from the IbAMP gene. IbAMP is a gene from the plant *Impatiens balsamina* which encodes a peculiar polyprotein precursor featuring a leader peptide and six consecutive antimicrobial peptides, each flanked by propeptides ranging from 16 to 28 amino acids in length (Tailor R.H. *et al.*, 1997, J. Biol. Chem. 272, 24480-24487). It is not known how and where processing of the IbAMP precursor occurs in its plant of origin. One of the internal propeptides from IbAMP was used to separate two distinct plant defensin coding regions, one originating from radish seed (RsAFP2, Terras F.R.G. *et al.*, 1992, J. Biol. Chem. 267, 15301-15309; Terras et al 1995 Plant Cell, 7, 573-588) and one from dahlia seed (DmAMP1, Osborn R.W. *et al.*, 1995, FEBS Lett. 368, 257-262). The other strategy was based on the use of C-terminal propeptides from either the DmAMP1 precursor or the AcAMP2 precursor (De Bolle M.F.C. *et al.*, 1993, Plant Mol. Biol. 22, 1187-1190). These C-terminal propeptides were chosen based on our previous observation that they apparently can be cleaved in transgenic tobacco plants without influencing extracellular deposition of the mature proteins to which they are connected in the precursor (R.W. Osborn and S.

Attenborough, personal communication; De Bolle M.F.C. *et al.*, 1996, Plant Mol. Biol. 31, 993-1008) implicating that such cleavage is performed by a protease present in the secretory pathway excluding the vacuole. To convert these C-terminal propeptides to internal propeptides, a subtilisin-like protease processing site was engineered at the C-terminal part of the propeptides. Subtilisin-like proteases are enzymes that specifically cleave at recognition sites of which the last two residues are basic (Barr, P.J., 1991, Cell 66, 1-3; Park C.M. *et al.*, 1994, Mol. Microbiol. 11, 155-164). Although subtilisin-like proteases are best documented in fungi (e.g. Kex2-like proteases) and higher animals (e.g. furins), recent evidence suggests that such enzymes are also present in plants (Kinal H. *et al.*, 1995, Plant Cell 7, 677-688; Tornero P. *et al.*, 1997, J. Biol. Chem. 272, 14412-14419), including Arabidopsis (Ribeiro A. *et al.*, 1995, Plant Cell 7, 785-794).

We have found that polyprotein precursors consisting of a leader peptide followed by two different plant defensins separated from each other by any of the above described internal propeptides can be processed in transgenic plants to release both plant defensins simultaneously. The cleavage does occur such that at least the major part of the plant defensins are deposited in the extracellular space. Hence processing of the precursor occurred either in the secretory pathway or in the extracellular space. The different propeptides shown to be cleaved in the transgenic plants do not reveal primary sequence homology. However, the sequences all appear to be rich in the small amino acids A, V, S and T and all contain dipeptidic sequences consisting of either two acidic residues, two basic residues or one acidic and one basic residue. Although propeptide cleavage in the examples shown in this invention did apparently not occur within vacuoles, internal propeptides from vacuolar proteins (e.g. 2S albumins) might also be used if vacuolar deposition of the proteins would be desirable. In the co-expression experiments described here two different plant defensins were used but it is predicted that similar results will be obtained when other types of proteins would be used or when more than two mature protein domains would be used in the polyprotein precursor structure.

Where it is desired to target the polyprotein to a particular cellular organelle along the secretory pathway a suitable targeting sequence may be added to one or more of the multiple protein encoding regions. For example, an endoplasmic reticulum targeting sequence such as that encoding KDEL may be added to the 3' end of one or more of the mature protein encoding regions, or a vacuolar targeting sequence (Chispeels and Raikhel 1992, Cell 68, 613-

616) can be added to the 3' or 5' end of one or more of the protein encoding regions. An example of the latter is the barley lectin carboxy-terminal propeptide which has been shown to destine heterologous proteins that are otherwise secreted to the vacuoles (Bednarek and Raikhel 1991, Plant Cell 3, 1195-1206; De Bolle et al, 1996 Plant Mol. Biol. 31, 993-1008).

At least 40% of the sequence of the linker propeptide for use in accordance with all aspects and methods of the invention as described herein preferably consists of stretches of either two to five consecutive hydrophobic residues selected from alanine, valine, isoleucine, methionine, leucine, phenylalanine, tryptophan and tyrosine or stretches of two to five hydrophilic residues selected from aspartic acid, glutamic acid, lysine, arginine, histidine, serine, threonine, glutamine and asparagine.

The said hydrophobic residues are preferably alanine, valine, leucine, methionine and/or isoleucine and the said hydrophilic residues are preferably aspartic acid, glutamic acid, lysine and/or arginine.

It is further preferred that the linker propeptide has within 7 residues of its N- or C-terminal cleavage site a sequence with two to five consecutive acidic residues, two to five basic residues or two to five consecutive intermixed acidic and basic residues.

It is especially preferred that at least 40% of the sequence of the linker propeptide for use in accordance with all aspects of the invention as described herein preferably consists of stretches of either two to five consecutive hydrophobic residues selected from alanine, valine, isoleucine, methionine, leucine, phenylalanine, tryptophan and tyrosine or stretches of two to five hydrophilic residues selected from aspartic acid, glutamic acid, lysine, arginine, histidine, serine, threonine, glutamine and asparagine and has within 7 residues of its N- or C- terminal cleavage site a sequence with two to five consecutive acidic residues, two to five basic residues or two to five consecutive intermixed acidic and basic residues.

The use of linker propeptides rich in the small amino acids A, V, S and T and containing dipeptidic sequences consisting of either two acidic residues, two basic residues or one acidic and one basic residue which on translation provides a cleavage site whereby the expressed polyprotein is post-translationally processed into the component protein molecules is also preferred.

As used herein the term 'rich' is used to denote that the residues A,V, S and T are present more frequently than would be expected based on a random distribution of amino acids.

It is further preferred that the linker propeptides have a dipeptidic sequence within seven amino acids from the N- and/or C- terminal ends thereof, the said dipeptidic sequences consisting of either two acidic residues, two basic residues or an acidic and a basic residue wherein said dipeptidic sequences may be the same or different at each terminus.

In a further preferred embodiment said dipeptidic sequences are selected from the following EE, ED and/or KK.

It is particularly desirable that the linker propeptide should hold the two (or more) protein domains sufficiently far apart so that they can fold appropriately and independently. It is further advantageous that the linker propeptide should not interact with any secondary structural element in the two proteins which it links and should therefore itself have no particular secondary structure or form a solitary secondary structure element such as an alpha helix.

In this and all other aspects and embodiments of the invention described herein the linker propeptide sequence providing the cleavage site is preferably isolatable from a plant protein, most preferably from the precursor of a plant antimicrobial protein such as a defensin, a hevein-type antimicrobial peptide (Broekaert et al 1997, Crit. Rev. Plant Sci. 16, 297-323) or from an antimicrobial peptide derived from the genus Impatiens and is most preferably a propeptide derived from the Ib-AMP gene, and/or the C-terminal propeptides from Dm-AMP1 and Ac-AMP2 the sequences of which are as described in Figure 2 herein. The Ib-AMP gene comprises five propeptide regions all of which are suitable for use in the present invention and which are described fully in Published International Patent Application WO 95/24486 at pages 29 and 40 to 42, the contents of which are incorporated herein by reference. All or part of the C-terminal propeptides derived from the Dm-AMP and Ac-AMP gene and all or part of any of the internal propeptides derived from the Ib-AMP gene may be used.

The use of the propeptide derived from the Ib-AMP gene as described in Figure 2 and also other propeptides derived from the Ib-AMP gene, the C-terminal propeptides from Dm-AMP1 and Ac-AMP2 as described in Figure 2 herein as cleavable linkers i.e. to provide a cleavable linkage site, are particularly preferred. Depending on the choice of propeptide it may be necessary to engineer an additional specific protease recognition site at either or both ends to facilitate cleavage of the sequence. Suitable specific protease recognition sites include for example, recognition sites for subtilisin -like proteases recognising either a dipeptidic sequence consisting of two basic residues; tetrapeptidic sequences consisting of a hydrophobic

residue, any residue, a basic residue and a basic residue or a tetrapeptidic sequence consisting of a basic residue, any residue, a basic residue and a basic residue. Subtilisin-like protease recognition sites are particularly preferred for use in the method of the invention.

The invention further provides the use of propeptides isolatable from plant derived proteins as cleavable linkers in polyprotein precursors synthesised via the secretory pathway in transgenic plants. The propeptides are preferably isolatable from the precursor of a plant defensin or a hevein-type antimicrobial peptide (Broekaert et al 1997, Crit. Rev. Plant Sci. 16, 297-323) or from an antimicrobial peptide derived from the genus Impatiens.

In a further aspect the invention provides the use of a propeptide wherein at least 40% of the sequence of the propeptide consists of stretches of either two to five consecutive hydrophobic residues selected from alanine, valine, isoleucine, methionine, leucine, phenylalanine, tryptophan and tyrosine or stretches of two to five hydrophilic residues selected from aspartic acid, glutamic acid, lysine, arginine, histidine, serine, threonine, glutamine and asparagine as a cleavable linker in polyprotein precursors synthesised via the secretory pathway in transgenic plants.

It is further preferred that the linker propeptide has within 7 residues of its N- or C-terminal cleavage site a sequence with two to five consecutive acidic residues, two to five basic residues or two to five consecutive intermixed acidic and basic residues.

It is especially preferred that at least 40% of the sequence of the linker propeptide consists of stretches of either two to five consecutive hydrophobic residues selected from alanine, valine, isoleucine, methionine, leucine, phenylalanine, tryptophan and tyrosine or stretches of two to five hydrophilic residues selected from aspartic acid, glutamic acid, lysine, arginine, histidine, serine, threonine, glutamine and asparagine and has within 7 residues of its N- or C- terminal cleavage site a sequence with two to five consecutive acidic residues, two to five basic residues or two to five consecutive intermixed acidic and basic residues.

In a further aspect the invention provides the use of a peptide sequence rich in the small amino acids A, V, S and T and containing dipeptidic sequences consisting of either two acidic residues, two basic residues or one acidic and one basic residue as a cleavable linker sequence wherein said sequence is isolatable from a plant defensin or a hevein-type antimicrobial protein.

The methods of the invention may be used to achieve efficient expression and secretion of any desired proteins and is particularly suitable for the expression of proteins which must

naturally be synthesised in the secretory pathway in order to be folded in a functional form such as, for example, glycosylated proteins and those with disulphide bridges. Additionally, it is extremely advantageous for proteins involved in the defence of a plant to attack by a pathogen to be secreted efficiently to the extracellular space since this is usually the initial site of pathogen attack and the present methods of the invention provide an effective means of delivering multiple proteins extracellularly.

The method of the invention is also particularly suitable for producing small peptides which may then be used for immunisation purposes i.e. the transgenic plant or a seed derived therefrom may be used directly as a foodstuff thereby passively immunising the recipient.

Examples of proteins which may be expressed according to the methods of the present invention include, for example, antifungal proteins described in Published International Patent Application Nos WO92/15691, WO92/21699, WO93/05153, WO93/04586, WO94/11511, WO95/04754, WO95/18229, WO95/24486, WO97/21814 and WO97/21815 including Rs-AFP1, Rs-AFP2, Dm-AMP1, Dm-AMP2, Hs-AFP1, Ah-AMP1, Ct-AMP1, Ct-AMP2, Bn-AFP1, Bn-AFP2, Br-AFP1, Br-AFP2, Sa-AFP1, Sa-AFP2, Cb-AMP1, Cb-AMP2, Ca-AMP1, Bm-AMP1, Ace-AMP1, Ac-AMP1, Ac-AMP2, Mj-AMP1, Mj-AMP2, Ib-AMP1, Ib-AMP2, Ib-AMP3, Ib-AMP4, PR-1 type proteins such as chitinases, glucanases such as beta1,3 and beta1,6 glucanases, chitin-binding lectins, zeamatins, osmotins, thionins and ribosome-inactivating proteins and peptides derived therefrom or antifungal proteins showing 85% sequence identity, preferably greater than 90% sequence identity, more preferably greater than 95% sequence identity with any of said proteins.

In the context of the present invention, two amino acid sequences with at least 85% similarity to each other have at least 85% similar (identical or conservatively replaced) amino acid residues in a like position when aligned optimally allowing for up to 3 gaps, with the proviso that in respect of the gaps a total of not more than 15 amino acid residues is affected. Likewise, two amino acid sequences with at least 90% similarity to each other have at least 90% identical or conservatively replaced amino acid residues in a like position when aligned optimally allowing for up to 3 gaps with the proviso that in respect of the gaps a total of not more than 15 amino acid residues is affected.

For the purpose of the present invention, a conservative amino acid is defined as one which does not alter the activity/function of the protein when compared with the unmodified protein. In particular, conservative replacements may be made between amino

acids within the following groups:

(i) Alanine, Serine, Glycine and Threonine

(ii) Glutamic acid and Aspartic acid

(iii) Arginine and Lysine

(iv) Isoleucine, Leucine, Valine and Methionine

(v) Phenylalanine, Tyrosine and Tryptophan

Sequence similarity may be calculated using sequence alignment algorithms known in the art such as, for example, the Clustal Method described by Myers and Miller (Comput. Appl. Biosci .4 11-17 (1988).) and Wilbur and Lipman (Proc. Natl. Acad. Sci. USA 80, 726-30 (1983) ) and the Watterman and Eggert method (The Journal of Molecular Biology (1987) 197, 723-728). The MegAlign Lipman Pearson one pair method (using default parameters) which may be obtained from DNAstar Inc, 1228 Selfpark Street, Madison, Wisconsin, 53715, USA as part of the Lasergene system may also be used.

The cleavable linkers are used to join two or more proteins of interest and provide cleavage sites whereby the polyprotein is post-translationally processed into the component protein molecules.

In a further aspect the invention provides a DNA construct comprising a DNA sequence comprising a promoter region operably linked to a plant derived signal sequence said signal sequence being operably linked to two or more protein encoding regions and a 3'-terminator region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide said propeptide providing a post-translational cleavage site.

The invention does not extend to the use of a DNA construct in the expression of multiple proteins in a transgenic plant where when said propeptide linker is derived from the Ib-AMP gene said protein encoding regions encode only three copies of Rs-AFP2.

In a preferred embodiment of this aspect the invention provides a DNA construct wherein said DNA sequence encoding said linker propeptide encodes an internal propeptide from the Ib-AMP gene, or the C-terminal propeptide from the Dm-AMP or from the Ac-AMP gene.

In a particularly preferred embodiment the invention provides a DNA construct as described above wherein when the DNA sequence encoding the linker propeptide is derived

from the Dm-AMP gene or from the Ac-AMP gene it additionally comprises one or more protease recognition sites at either or both ends thereof.

In a further aspect the invention provides a DNA construct comprising a DNA sequence comprising a promoter region operably linked to two or more protein encoding regions and a 3' terminator-region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide encoding the C-terminal propeptide from the Dm-AMP gene or the from the Ac-AMP gene said propeptide providing a post-translational cleavage site.

In a particularly preferred embodiment the invention provides a DNA construct as described above wherein the DNA sequence encoding the linker propeptide from Dm-AMP or Ac-AMP additionally comprises one or more protease recognition sites at either or both ends thereof.

In a yet further aspect the invention provides a transgenic plant transformed with a DNA construct according to any of the above aspects of the invention.

In a further aspect the invention provides a transgenic plant transformed with a DNA sequence comprising a promoter region operably linked to a signal sequence said signal sequence being operably linked to two or more protein encoding regions and a 3'-terminator region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide which on translation provides a cleavage site.

The invention does not extend to a transgenic plant where when the protein encoding regions are separated by a linker propeptide derived from the Impatiens gene they encode only three copies of the Rs-AFP2 protein.

In a preferred embodiment of this aspect at least 40% of the sequence of the said linker propeptide consists of stretches of either two to five consecutive hydrophobic residues selected from alanine, valine, isoleucine, methionine, leucine, phenylalanine, tryptophan and tyrosine or stretches of two to five hydrophilic residues selected from aspartic acid, glutamic acid, lysine, arginine, histidine, serine, threonine, glutamine and asparagine.

The said hydrophobic residues are preferably alanine, valine, leucine, methionine and/or isoleucine and the said hydrophilic residues are preferably aspartic acid, glutamic acid, lysine and/or arginine.

It is further preferred that the linker propeptide has within 7 residues of its N- or C-terminal cleavage site a sequence with two to five consecutive acidic residues, two to five basic residues or two to five consecutive intermixed acidic and basic residues.

It is especially preferred that at least 40% of the sequence of the linker propeptide consists of stretches of either two to five consecutive hydrophobic residues selected from alanine, valine, isoleucine, methionine, leucine, phenylalanine, tryptophan and tyrosine or stretches of two to five hydrophilic residues selected from aspartic acid, glutamic acid, lysine, arginine, histidine, serine, threonine, glutamine and asparagine and has within 7 residues of its N- or C- terminal cleavage site a sequence with two to five consecutive acidic residues, two to five basic residues or two to five consecutive intermixed acidic and basic residues.

In a further preferred embodiment of this aspect of the invention the DNA sequence providing the cleavage site encodes a peptide sequence rich in the small amino acids A, V, S and T and containing dipeptidic sequences consisting of either two acidic residues, two basic residues or one acidic and one basic residue.

In a particularly preferred embodiment of this aspect of the invention the DNA sequence providing the cleavage site encodes a propeptide derived from the Ib-AMP gene such as for example that described in Figure 2 or the C-terminal propeptides from Dm-AMP1 and Ac-AMP2 as described in Figure 2 which may optionally be engineered to include a further DNA sequence encoding a subtilisin-like protease recognition site.

In a further aspect the invention provides a vector comprising a DNA construct as described above.

Unexpectedly, expression levels of plant defensins in plants transformed with a polyprotein precursor construct were found to be much higher compared to those in plants transformed with single plant defensin constructs. Hence, the processing system described here can be used not only to co-express two or more different proteins, but also to obtain higher expression levels of a protein, particularly of small proteins. The reason for the observed stimulatory effect on translational efficiency is currently unclear. It might be due to an effect of mRNA length or length of primary translation product on translational efficiency.

In a further aspect the invention therefore provides a method of improving expression levels of a protein in a transgenic plant comprising inserting into the genome of said plant a DNA sequence comprising a promoter region operably linked to two or more protein encoding regions and a 3'-terminator region wherein said protein encoding regions are separated from

each other by a DNA sequence coding for a linker propeptide said propeptide providing a cleavage site whereby the expressed polyprotein is post-translationally processed into the component protein molecules.

In a further preferred embodiment of this aspect there is provided a method of improving expression levels of a protein in a transgenic plant comprising inserting into the genome of said plant a DNA sequence comprising a promoter region operably linked to a signal sequence said signal sequence being operably linked to two or more protein encoding regions and a 3'-terminator region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide said propeptide providing a cleavage site whereby the expressed polyprotein is post-translationally processed into the component protein molecules.

As will be readily apparent to a man skilled in the art the sequence of the individual components of the DNA sequence i.e. the signal sequence, promoter sequence, linker sequence, protein sequence(s), terminator sequence for use in the methods according to the invention may be predicted from its known amino acid sequence and DNA encoding the protein may be manufactured using a standard nucleic acid synthesiser. Alternatively, DNA encoding the components of the invention may be produced by appropriate isolation from natural sources.

The invention is further illustrated with reference to the following non-limiting examples and figures in which

Figure 1: shows nucleotide sequence and corresponding amino acid sequence of coding region of the DmAMP1 gene. The amino acids corresponding to mature DmAMP1 are underlined. The nucleotides corresponding to the intron are double underlined.

Figure 2: shows schematic representation of the coding regions from the vector constructs. Amino acids sequences below the internal propeptides represent the propeptide sequences from which the linker propeptides were derived.

Figure 3: shows schematic representation of plant transformation vector pFAJ3105

Figure 4: shows schematic representation of plant transformation vector pFAJ3106

Figure 5: shows schematic representation of plant transformation vector pFAJ3107

Figure 6: shows schematic representation of plant transformation vector pFAJ3108

Figure 7: shows schematic representation of plant transformation vector pFAJ3109

Figure 8: shows nucleotide sequence and corresponding amino acid sequence of the open reading frame of the region comprised between the *Nco*I and *Sac*I sites of plasmid pFAJ3105. The amino acids corresponding to mature DmAMP1 and mature RsAFP2 are underlined and double-underlined, respectively.

Figure 9: shows nucleotide sequence and corresponding amino acid sequence of the open reading frame of the region comprised between the *Nco*I and *Sac*I sites of plasmid pFAJ3106. The amino acids corresponding to mature DmAMP1 and mature RsAFP2 are underlined and double-underlined, respectively.

Figure 10: shows nucleotide sequence and corresponding amino acid sequence of the open reading frame of the region comprised between the *Nco*I and *Sac*I sites of plasmid pFAJ3107. The amino acids corresponding to mature DmAMP1 and mature RsAFP2 are underlined and double-underlined, respectively.

Figure 11: shows nucleotide sequence and corresponding amino acid sequence of the open reading frame of the region comprised between the *Nco*I and *Sac*I sites of plasmid pFAJ3108. The amino acids corresponding to mature DmAMP1 and mature RsAFP2 are underlined and double-underlined, respectively.

Figure 12: shows nucleotide sequence and corresponding amino acid sequence of the open reading frame of the region comprised between the *Nco*I and *Sac*I sites of plasmid pFAJ3109. The amino acids corresponding to mature DmAMP1 are underlined.

Figure 13: shows the Dm-AMP1 expression levels (as % of total soluble protein) of a series of transgenic individual plants transformed with construct pFAJ3105 and a series of transgenic individuals transformed with construct pFAJ3109.

Figure 14: shows RP-HPLC analysis on a C8-silica column of crude extracts from leaves transformed with construct pFAJ3105 (A) or pFAJ3106 (B). Extracts were prepared as described in Materials and Methods. The column was eluted with a gradient of acetonitrile in 0.1 % TFA (0-35 min. 15 % - 50 % acetonitrile in 0.1 % TFA). The eluate was monitored on-line for measurement of the absorbance at 214 nm (top trace), fractionated, and subjected to Elisa assays for DmAMP1 (lower bar graph, black bars) and RsAFP2 (lower bar graph, white bars). The elution position of authentic DmAMP1 and RsAFP2 are indicated with arrows on the $A_{214}$ chromatograms.

## EXAMPLES

## MATERIALS AND METHODS

### Cloning of DmAMP1 cDNA and DmAMP1 gene

Cloning procedures and polymerase chain reaction (PCR) procedures were performed following standard protocols (Sambrook *et al.*, 1989, Molecular Cloning: a laboratory manual, 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). A cDNA library was constructed from near-dry seeds collected from flowers of *Dahlia merckii*. Total RNA was purified from the seeds using the method of Jepson I. *et al.* (1991, Plant Mol. Biol. Reporter 9, 131-138). 0.6 mg of total RNA was obtained from 2 g of *D. merckii* seed. PolyATract magnetic beads (Promega) were used to isolate approximately 2 µg poly-A+ RNA from 0.2 mg of total RNA.

The poly-A+ RNA was used to construct a cDNA library using a ZAP-cDNA synthesis kit (Stratagene). Following first and second strand synthesis, cDNAs were ligated with vector DNA. After phage assembly using Gigapack Gold (Stratagene) packaging extracts, approximately 1 x $10^5$ plaque forming units (pfu) were obtained.

Using oligonucleotides AFP-5 (5'-TG(T,C)GANAANGCN(A,T)(G,C)NAA(A,G)ACNTGG) based on the N-terminal sequence CEKASKTW of DmAMP1, Osborn R.W. *et al.*, 1995, FEBS Lett. 368, 257-262) and AFP-3EX (5'-CA(A,G)TT(A,G)AANTANCANAAA(A,G) CACAT) based on the C-terminal sequence MCFCYFNC of DmAMP1) and genomic DNA isolated from *D. merckii* leaves, a 144 bp PCR product was produced and isolated from an agarose gel. The PCR product was cloned into pBluescript. The insert of 10 transformants were sequenced. The sequences represented 3 closely homologous DmAMP1-like genes one of which, PCR clone 4, encoded the observed mature DmAMP1. The 144 bp PCR product mixture labelled with $^{32}$-P CTP was used to probe Hybond N (Amersham) filter lifts made from plates containing a total of 6 x $10^4$ pfu of the cDNA library. Thirty potentially positive signals were observed. 22 plaques were picked and taken through two further rounds of screening. After in vivo excision 13 clones were characterised by DNA sequencing. Four classes of DmAMP related peptides were encoded by the 13 cDNA clones. Three versions of the DmAMP mature protein region were represented in the four classes. One of the classes (Dm2.5 type) contained a mature protein region which may correspond to DmAMP2 (Osborn R.W. *et al.*, 1995, FEBS Lett. 368, 257-262). None of the cDNAs encoded a mature protein region equivalent to the observed mature DmAMP1 peptide sequence.

Using the sequence of PCR clone 4 (above) and information from the N- and C-terminal ends of the peptides deduced from cDNA sequences, two pairs of oligonucleotides were designed for amplification of a gene encoding DmAMP1. Genomic DNA from *D. merckii* was used in a PCR reaction with oligonucleotides MATAFP-5P (5'-ATGGC(C,G)AAN(A,C)(A,G)NTC (A,G)GTTGCNTT) and MATAFP-5 (5'- AAACACATGTGTTTCCCATT), the PCR product was cloned into pBluescript and clones were sequenced. A clone containing the 5' half of a DmAMP1 gene was identified. Genomic DNA from *D. merckii* was used in a PCR reaction with MATAFP-3 (5'- AGCGTGTCATGTGCGTAAT) and DM25MAT-3 (5'- TAAAGA AACCGACCCTTTCACGG), the PCR product was cloned into pBluescript and clones were sequenced. A clone containing the 3' half of a DmAMP1 gene was identified. The 5' and 3' sections of the mature gene were combined to assemble the sequence of the coding region of the DmAMP1 gene (Figure 1).

The DmAMP1 gene encodes a precursor with a 28 amino acids leader peptide, a 50 amino acids mature protein and a 40 amino acids C-terminal propeptide. The open reading frame is interrupted by a 92 bp intron located within the leader peptide region.

To eliminate the intron from the DmAMP1 gene sequence and to allow cloning of the DmAMP1 encoding region, either with or without the C-terminal propeptide region, into an expression cassette vector, two PCR reactions were carried out with respectively the primer sets DMVEC-3 (5'- ATGCATCCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTT CTGATCCTTTTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATG CGAGAAA) and DMVEC-2 (5'- AAACCGACCGAGCTCACGGATGTTCAACGTTTGGA AC), and DMVEC-3 and DMVEC4 (5'- AGCAAGCTTTTCGGGAGCTCAACAATTGA AGTAA). DMVEC-3 primes at the top strand of the DmAMP1 gene, corresponds to the leader peptide region without the intron and introduces an *Nco*I site at the translation start. DMVEC-2 primes at the bottom strand of the DmAMP1 gene at the 3'-end of the C-terminal propeptide region and introduces a *Sac*I site behind the translation stop codon. DMVEC-4 primes at the bottom strand of the DMAMP1 gene at the 3' end of the mature protein region, fuses a stop codon behind this region and introduces a *Sac*I site behind the stop codon. Both PCR products were cut with *Nco*I and *Sac*I which cleaved the PCR products in two fragments due to an internal *Nco*I site in the mature protein region. The resulting *Nco*I-*Sac*I and *Nco*I-*Nco*I fragments were cloned sequentially in plasmid pMJB1. pMJB1 is an expression cassette vector containing in sequence a *Hin*dIII site, the enhanced cauliflower mosaic 35S RNA

(CaMV35S) promoter (Kay R. *et al.*, 1987, Science 236, 1299-1302), a *Xho*I site, the 5' untranslated leader sequence of tobacco mosaic virus (TMV) (Gallie D.R. and Walbot V., 1992, Nucl. Ac. Res. 20, 4631-4638) a polylinker including *Nco*I, *Sma*I, *Kpn*I and *Sac*I sites, the 3' untranslated terminator region of the *Agrobacterium tumefaciens* nopaline synthase gene (Bevan M.W. *et al.*, 1983, Nature 304, 184-187) and an *Eco*RI site. The resulting plasmids were termed pDMAMPE (leader peptide region, mature protein region and C-terminal propeptide region) and pDMAMPD (leader peptide region and mature protein region), respectively. The coding regions were verified by DNA sequencing.

**Constructions of plant transformation vectors**

Schematic representations of the plant transformation vectors used in this work, pFAJ3105, pFAJ3106, pFAJ3107, pFAJ3108 and pFAJ3109, are shown in figures 3 till 7, respectively. The nucleotide sequences comprised between the *Xho*I and *Sac*I sites of these plasmids, which encompass the regions encoding antimicrobial proteins, are presented in Figures 8 till 13. The regions comprised between the *Xho*I and *Sac*I sites of plasmid pFAJ3105 (shown in Figure 8) was constructed following the two-step recombinant PCR protocol of Pont-Kindom G.A.D. (1994, Biotechniques 16, 1010-1011). Primers OWB175 (5'AGGAAGTTCATTTCATTTGG) and OWB278 (5'-GCCTTTGGCACAACTTCTGT CCTGGCTCCACGTCCTCTGGGGTAGCCACCTCGTCAGCAGCGTTGGAACAATTGA AGTAACAGAAACAC) were used in a first PCR reaction with plasmid pDMAMPE (see above) as a template. The second PCR reaction was done using as a template plasmid pFRG4 (Terras F.R.G. *et al.*, 1995, Plant Cell 7, 573-588) and as primers a mixture of the PCR product of the first PCR reaction, primer OWB175 and primer OWB172 (5'TTAGAGCTCCTATTAACAAGGAAAGTAGC, *Sac*I site underlined). The resulting PCR product was digested with *Xho*I and *Sac*I and cloned into the expression cassette vector pMJB1 (see above). The expression cassette in the resulting plasmid, called pFAJ3099, was digested with *Hin*dIII (flanking the 5' end of the CaMV35S promoter) and *Eco*RI (flanking the 3' end of the nopaline synthase terminator) and cloned in the corresponding sites of the plant transformation vector pGPTVbar (Becker D. *et al.*, 1992, Plant Mol. Biol. 20, 1195-1197) to yield plasmid pFAJ3105.

Plasmids pFAJ3106, pFAJ3107 and pFAJ3108 were constructed analogously except that primer OWB278 in the first PCR reaction was replaced by the following primers, respectively:

OWB279 (5'-GCCTTTGGCACAACTTCTGCCTCTTTCCGATGAGTTGTTCGGCTTT AAGTTTGTC); OWB303 (5'-GCCTTTGGCACAACTTCTGCCTCTTTCCG ATCGGATGTTCAACGTTTGGAACC); OWB304 (5'-GCCTTTGGCACAACTTCTGCCT CTTTCCGATAGTTTTGGTGGCAGCAACATCAGCTTGGTGATCCACAGTAGTACTGG CACAATTGAAGTAACAGAAACAC).

Plasmid pFAJ3109 was constructed by cloning the *Hin*dIII-*Eco*RI fragment of plasmid pDMAMPD (see above) into the corresponding sites of plant transformation vector pGPTVbar (see above).

**Plant transformation**

*Arabidopsis thaliana* ecotype Columbia-O was transformed using recombinant *Agrobacterium tumefaciens* by the inflorescence infiltration method of Bechtold N. *et al.* (1993, C.R. Acad. Sci. 316, 1194-1199). Transformants were selected on a sand/perlite mixture subirrigated with water containing the herbicide Basta (Agrevo) at a final concentration of 5 mg/l for the active ingredient phosphinothricin.

**Elisa assays and protein assays**

Antisera were raised in rabbits injected with either RsAFP2 (purified as described in Terras F.R.G. *et al.*, 1992, J. Biol. Chem. 267, 15301-15309) or DmAMP1 (purified as in Osborn R.W. *et al.*, 1995, FEBS Lett. 368, 257-262). ELISA assays were set up as competitive type assays essentially as described by Penninckx I.A.M.A. *et al.* (1996, Plant Cell 8, 2309-2323). Coating of the ELISA microtiter plates was done with 50 ng/ml RsAFP2 or DmAMP1 in coating buffer. Primary antisera were used as 1000- and 2000-fold diluted solutions (DmAMP1 and RsAFP2, respectively) in 3 % (w/v) gelatin in PBS containing 0.05 % (v/v) Tween 20.

Total protein content was determined according to Bradford (1976, Anal. Biochem. 72, 248-254) using bovine serum albumin as a standard.

**Rough separation of proteins processed from polyprotein precursors**

Arabidopsis leaves were homogenized under liquid nitrogen and extracted with a buffer consisting of 10 mM $NaH_2PO_4$, 15 mM $Na_2HPO_4$, 100 mM KCl, 1.5 M NaCl. The homogenate was heated for 10 min at 85°C and cooled down on ice. The heat-treated extract was centrifuged for 15 min at 15 000 x *g* and was injected on a reserved phase high pressure liquid chromatography column (RP-HPLC) consisting of C8 silica (0,46 cm x 25 cm; Rainin) equilibrated with 0.1 % (v/v) trifluoroacetic acid (TFA). The column was eluted at 1 ml/min

in a linear gradient in 35 min from 15 % to 50 % (v/v) acetonitrile in 0.1 % (v/v) TFA. The eluate was monitored for absorbance at 214 nm, collected as 1 ml fractions, evaporated and finally redissolved in water. The fractions were tested by ELISA assays.

**Preparation of extracellular fluid and intracellular extract**

Intercellular fluid was collected from Arabidopsis leaves by immersing the leaves in a beaker containing extraction buffer (10 mM $NaH_2PO_4$, 15 mM $Na_2HPO_4$, 100 mM KCl, 1.5 M NaCl). The beaker with the leaves was placed in a vacuum chamber and subjected to six consecutive rounds of vacuum for 2 min followed by abrupt release of vacuum. The infiltrated leaves were gently placed in a centrifuge tube on a grid separated from the tube bottom. The intercellular fluid was collected from the bottom after centrifugation of the tubes for 15 min at 1800 x *g*. The leaves were resubjected to a second round of vacuum infiltration and centrifugation and the resulting (extracellular) fluid was combined with that obtained after the first vacuum infiltration. After this step the leaves were extracted in a Phastprep (BIO101/Savant) reciprocal shaker and the extract clarified by centrifugation (10 min at 10,000 x *g*) and the resulting supernatant considered as the intracellular extract.

**RESULTS**

**Characterization of transgenic plants and expression analysis**

To explore the possibility of expressing polyprotein precursor genes in plants, four different plant transformation vectors were made with the aim to co-express two different cysteine-rich plant defensins with antifungal properties, namely RsAFP2 and DmAMP1. The polyprotein precursor regions of these constructs all featured a leader peptide region derived from the DmAMP1 cDNA, the mature protein domain of DmAMP1, an internal propeptide region, and the mature protein domain of RsAFP2. The four constructs differed only in the internal propeptides (Figure 2):

- construct 3105 has one of the IbAMP internal propeptides as a propeptide separating DmAMP1 and RsAFP2.
- construct 3106 has a propeptide consisting of a part of the DmAMP1 propeptide and a putative subtilisin-like protease processing site (IGKR) at its C-terminus.
- construct 3107 is identical to construct 3106 except that the entire DmAMP1 propeptide was taken.
- construct 3108 has a propeptide consisting of the AcAMP2 propeptide and a putative subtilisin-like protease processing site (IGKR) at its C-terminus.

The rationale behind constructs 3106, 3107 and 3108 is based on our observations that the C-terminal propeptides of AcAMP2 and DmAMP1 are cleaved off at their N-terminus when expressed as AcAMP2- and DmAMP1-preproproteins in tobacco, respectively, while this processing event does not detract the mature proteins from being sorted to the apoplast (De Bolle *et al.*, 1996, Plant Mol. Biol. 31, 993-1008; R.W. Osborn and S. Attenborough, personal communication). This infers that the processing enzymes are either in the secretory pathway or in the apoplast. On the other hand, C-terminal cleavage of the internal propeptide in these constructs should be executed by a subtilisin-like protease, a member of which in yeast (Kex2) is known to occur in the Golgi apparatus (Wilcox C.A. and Fuller R.S., 1991, J. Cell. Biol. 115, 297), while a member in tomato occurs in the apoplast (Tornero P. *et al.*, 1997, J. Biol. Chem. 272, 14412-14419). Proteins deposited in the apoplast, the preferred deposition site for antimicrobial proteins engineered in transgenic plants (Jongedijk E. *et al.*, 1995, Euphytica 85, 173-180; De Bolle *et al.*, 1996, Plant Mol. Biol. 31, 993-1008) are normally synthesized via the secretory pathway, encompassing the Golgi apparatus.

A construct was also made for expression of only DmAMP1 (construct 3109, figure 7). Expression levels of DmAMP1 and RsAFP2 were analysed in leaves taken from a series of T1 transgenic Arabidopsis plants resulting from transformation with the constructs described above. The results of the expression analyses based on Elisa assays are presented in Table 1. Most of the tested lines transformed with the polyprotein constructs 3105, 3106, 3107 and 3108 clearly expressed both DmAMP1-CRPs (DmAMP1-crossreactive proteins) and RsAFP2-CRPs (Rs-AFP2-crossreactive proteins). There was generally a good correlation between DmAMP1-CRP and RsAFP2-CRP levels. However, the RsAFP2-CRP levels were generally 2 to 5-fold lower than the DmAMP1-CRP levels. The Elisa assays for measuring the RsAFP2-CRPs in the extracts are, however, less reliable than those for the Dm-AMP1-CRPs. In Rs-AFP2 Elisa assays, dilutions of extracts of transgenic plants yielded dose-response curves that deviated from those obtained for dilutions of standard solutions containing authentic Rs-AFP2, indicating that the majority of the Rs-AFP2 -CRPs in the extracts were imunologically not identical to RsAFP2 itself. Deviations from RsAFP2 standard dose-response curves were much more pronounced for extracts from plants transformed with constructs 3106, 3107, and 3108 than for those of plants transformed with 3105. None of the extracts showed deviations from Dm-AMP1 standards in dose response curves in Dm-AMP1 Elisa assays. The DmAMP-CRP levels in the lines transformed with the polyprotein constructs 3105, 3106,

3107 or 3108 were generally much higher compared to those in the line transformed with the single protein construct 3109. This is also illustrated in Figure 13 where DmAMP1-CRP expression levels are compared for plants transformed with the polyprotein construct 3105 and plants transformed with the single protein construct 3109. Expression levels as high as 4% of total protein (e.g. DmAMP1-CRP level in lines 3105-15 and 3105 -18, see table 1) have so far never been reported in the literature for a peptide expressed in transgenic plants. Hence, the use of polyprotein constructs appears to result in markedly enhanced expression, which is an unexpected finding.

**Rough separation of proteins processed from polyprotein precursors**

A transgenic line was selected among each of the populations transformed with either construct 3105 (line 1) or 3106 (line 2) and the selected lines were further bred to obtain plants homozygous for the transgenes. In order to analyse whether DmAMP1 and RsAFP2 were correctly processed in these lines, extracts from the plants were prepared as described in Materials and Methods and separated by RP-HPLC on a C8-silica column. Fractions were collected and assessed for presence of compounds cross-reacting with antibodies raised against either DmAMP1 or RsAFP2 using Elisa assays.

As shown in figure 15, DmAMP1- CRPs eluted at a position identical or very close to that of authentic DmAMP1 in the line transformed with construct 3105 as well as in that transformed with construct 3106. Likewise, RsAFP2-CRPs were detected in both the construct 3105 and 3106 lines at an elution position identical or very close to that of authentic RsAFP2. None of the fractions reacted with both the anti-DmAMP1 and anti-RsAFP2 antibodies, indicating that an uncleaved fusion protein was not present in the extracts. No cross-reacting compounds were observed in a non-transformed line.

It is concluded that the primary translation products of the transcription units of construct 3105 (IbAMP internal propeptide as linker peptide) and construct 3106 (partial DmAMP1 C-terminal propeptide with subtilisin-like protease site as a linker peptide) are somehow processed to yield separate DmAMP1-CRPs and RsAFP2-CRPs that appear to be identical or very closely related to DmAMP1 and RsAFP2, respectively, based on their chromatographic behavior.

**Analysis of the subcellular location of coexpressed plant defensins**

In order to determine whether the coexpressed plant defensins are either secreted extracellularly or deposited intracellularly, extracellular fluid and intracellular extract fractions

were obtained from leaves of homozygous transgenic Arabidopsis lines transformed with either constructs 3105 (line 2), 3106 (line 2) or 3108 (line 12). The cytosolic enzyme glucose-6-phosphate dehydrogenase was used as a marker to detect contamination of the extracellular fluid fraction with intracellular components. As shown in Table 2, glucose-6-phosphate dehydrogenase was partitioned in a ratio of about 80/20 between intracellular extract fractions and extracellular fluid fractions. In contrast, the majority of DmAMP1-CRP and RsAFP2-CRP content in all transgenic plants tested was found in the extracellular fluid fractions. These results indicate that both plant defensins released from the polyprotein precursors are deposited primarily in the apoplast. Hence, all processing steps that result in cleavage of the polyprotein structure must occur either in the apoplast or along the secretory pathway i.e. in the endoplasmic reticulum, the Golgi apparatus or in vesicles trafficking between Golgi and apoplast.

**Table 1:Expression levels of Dm-AMP1 and Rs-AFP2 in transgenic Arabidopsis lines**

| construct | line | Expression level of Dm-AMP1 (%) | expression level of Rs-AFP2 (%) |
|---|---|---|---|
| 3105 | 1 | 0,77 | 0,29 |
| | 2 | 1,13 | 0,22 |
| | 3 | 0,48 | 0,20 |
| | 4 | 0,005 | <0,001 |
| | 5 | 0,36 | 0,05 |
| | 6 | 0,99 | 0,25 |
| | 7 | 0,60 | 0,09 |
| | 8 | 0,13 | <0,001 |
| | 9 | 0,25 | 0,08 |
| | 10 | 4,15 | 0,85 |
| | 11 | 1,35 | 0,35 |
| | 12 | 0,24 | 0,07 |
| | 13 | 4,43 | 0,91 |
| | 14 | 1,18 | 0,24 |
| | 15 | 0,68 | 0,17 |
| | 16 | 0,49 | 0,07 |
| 3106 | 1 | 0,10 | 0,001 |
| | 2 | 1,82 | 0,008 |
| | 3 | 0,68 | 0,20 |
| | 4 | 1,15 | 0,38 |
| | 5 | 0,20 | 0,10 |
| | 6 | 0,10 | 0,05 |
| | 7 | 0,40 | 0,17 |
| | 8 | 2,64 | 0,50 |
| | 9 | 0,40 | 0,15 |
| | 10 | 0,21 | 0,07 |
| | 11 | 0,06 | 0,03 |
| | 12 | 0,24 | 0,09 |
| 3107 | 1 | 0,04 | 0,04 |
| | 2 | 0,75 | 0,42 |
| | 3 | 0,14 | 0,13 |
| | 4 | 0,01 | 0,01 |
| | 5 | 0,27 | 0,29 |
| 3108 | 1 | 0,47 | 0,10 |
| | 2 | 3,00 | 0,53 |
| | 3 | 0,91 | 0,24 |
| | 4 | 2,04 | 0,22 |
| | 5 | 0,17 | 0,04 |
| | 6 | 0,55 | 0,05 |
| | 7 | 0,16 | 0,11 |
| | 8 | 0,05 | 0,02 |
| | 9 | 0,45 | 0,02 |
| 3109 | 1 | 0,19 | nd |
| | 2 | 0,05 | nd |
| | 3 | 0,02 | nd |
| | 4 | 0,20 | nd |
| | 5 | 0,10 | nd |
| | 6 | 0,06 | nd |
| | 7 | 0,07 | nd |
| | 8 | 0,003 | nd |
| | 9 | 0,18 | nd |

Table 2: Relative abundance of glucose-6-phosphate dehydrogenase activity (GPD), DmAMP1 and RsAFP2 in the extracellular fluid (EF) and intracellular extract (IE) fractions obtained from transgenic Arabidopsis plants.

| Construct | Relative abundance[1] (%) of | | | | | |
|---|---|---|---|---|---|---|
| | GPD | | DmAMP1 | | RsAFP2 | |
| | EF | IE | EF | IE | EF | IE |
| pFAJ3105 | 17 | 83 | 93 | 7 | 92 | 8 |
| pFAJ3106 | 17 | 83 | 94 | 6 | 60 | 40 |
| pFAJ3108 | 20 | 80 | 98 | 2 | 75 | 25 |

[1]Relative abundance is expressed as %.of the sum of the contents in the EF and IE fractions.

pFAJ3105

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

TTCAATTGTTCCAACGCTGCTGACGAGGTGGCTACCCCAGAGGACGTGGAGCCAGGACAG

F N C **S N A A D E V A T P E D V E P G** Q

AAGTTGTGCCAAAGGCCAAGTGGGACATGGTCAGGAGTCTGTGGAAACAATAACGCATGC

K L C Q R P S G T W S G V C G N N N A C

AAGAATCAGTGCATTAGACTTGAGAAAGCACGACATGGATCTTGCAACTATGTCTTCCCA

K N Q C I R L E K A R H G S C N Y V F P

SacI

GCTCACAAGTGTATCTGCTACTTTCCTTGTTAATAGGAGCTC

A H K C I C Y F P C - -

pFAJ3106

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

TTCAATTGTAAAAAAGCCGAAAAGCTTGCTCAAGACAAACTTAAAGCCGAACAACTCATC

F N C **K K A E K L A Q D K L K A E Q L I**

GGAAAGAGGCAGAAGTTGTGCCAAAGGCCAAGTGGGACATGGTCAGGAGTCTGTGGAAAC

**G K R** Q K L C Q R P S G T W S G V C G N

AATAACGCATGCAAGAATCAGTGCATTAGACTTGAGAAAGCACGACATGGATCTTGCAAC

N N A C K N Q C I R L E K A R H G S C N

SacI

TATGTCTTCCCAGCTCACAAGTGTATCTGCTACTTTCCTTGTTAATAGGAGCTC

Y V F P A H K C I C Y F P C - -

pFAJ3107

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

TTCAATTGTAAAAAAGCCGAAAAGCTTGCTCAAGACAAACTTAAAGCCGAACAACTCGCT

F N C K K A E K L A Q D K L K A E Q L A

CAAGACAAACTTAATGCCCAAAAGCTTGACCGTGATGCCAAGAAAGTGGTTCCAAACGTT

Q D K L N A Q K L D R D A K K V V P N V

GAACATCCGATCGGAAAGAGGCAGAAGTTGTGCCAAAGGCCAAGTGGGACATGGTCAGGA

E H P I G K R Q K L C Q R P S G T W S G

GTCTGTGGAAACAATAACGCATGCAAGAATCAGTGCATTAGACTTGAGAAAGCACGACAT

V C G N N N A C K N Q C I R L E K A R H

GGATCTTGCAACTATGTCTTCCCAGCTCACAAGTGTATCTGCTACTTTCCTTGTTAATAG

G S C N Y V F P A H K C I C Y F P C - -

SacI

GAGCTC

pFAJ3108

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

TTCAATTGTGCCAGTACTACTGTGGATCACCAAGCTGATGTTGCTGCCACCAAAACTATC

F N C **A S T T V D H Q A D V A A T K T I**

GGAAAGAGGCAGAAGTTGTGCCAAAGGCCAAGTGGGACATGGTCAGGAGTCTGTGGAAAC

**G K R** Q K L C Q R P S G T W S G V C G N

AATAACGCATGCAAGAATCAGTGCATTAGACTTGAGAAAGCACGACATGGATCTTGCAAC

N N A C K N Q C I R L E K A R H G S C N

SacI

TATGTCTTCCCAGCTCACAAGTGTATCTGCTACTTTCCTTGTTAATAGGAGCTC

Y V F P A H K C I C Y F P C - -

pFAJ3109

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAATGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

SacI

TTCAATTGTTGAGCTC

F N C -

CLAIMS

1. A method for the expression of multiple proteins in a transgenic plant comprising inserting into the genome of said plant a DNA sequence comprising a promoter region operably linked to a signal sequence said signal sequence being operably linked to two or more protein encoding regions and a 3'-terminator region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide said propeptide providing a cleavage site whereby the expressed polyprotein is post-translationally processed into the component protein molecules.

2. A method of improving expression levels of one or more proteins in a transgenic plant comprising inserting into the genome of said plant a DNA sequence comprising a promoter region operably linked to two or more protein encoding regions and a 3'-terminator region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide said propeptide providing a cleavage site whereby the expressed polyprotein is post-translationally processed into the component protein molecules.

3. A method for improving expression levels of one or more proteins in a transgenic plant according to claim 2 wherein said promoter region is operably linked to a signal sequence said signal sequence being operably linked to two or more protein encoding regions and a 3'-terminator region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide said propeptide providing a cleavage site whereby the expressed polyprotein is post-translationally processed into the component protein molecules.

4. A method according to any of the preceding claims wherein at least 40% of the sequence of said linker propeptide consists of stretches of either two to five consecutive hydrophobic residues selected from alanine, valine, isoleucine, methionine, leucine, phenylalanine, tryptophan and tyrosine or stretches of two to five hydrophilic residues selected from aspartic acid, glutamic acid, lysine, arginine, histidine, serine, threonine, glutamine and asparagine.

5. A method according to any of the preceding claims wherein said linker propeptide has within 7 residues of its N- or C- terminal cleavage site a sequence with two to five consecutive acidic residues, two to five basic residues or two to five consecutive intermixed acidic and basic residues.

6. A method according to any of the preceding claims wherein the DNA sequence encoding said linker propeptide encodes a propeptide isolatable from a plant protein.

7. A method according to claim 6 wherein the plant protein is a precursor of a plant defensin, a hevein-type antimicrobial protein or is an antimicrobial protein derived from the genus Impatiens.

8. A method according to claim 7 wherein the propeptide is isolatable from the Ib-AMP precursor, the Ib-AMP precursor as described in Figure 2, or the C-terminal propeptides from Dm-AMP1 or Ac-AMP2 as described in Figure 2.

9. A method according to any of the preceeding claims wherein the linker propeptide has a protease processing site engineered at either or both ends thereof.

10. A method according to claim 9 wherein the protease processing site is a subtilisin -like protease processing site.

11. A method according to any of claims 1 and 3 to 5 wherein the signal sequence is derived from a plant defensin gene.

12. A method according to any of the preceding claims wherein one or more of the multiple proteins is a defense protein.

13. Use of propeptides derived from plant derived proteins as cleavable linkers in polyprotein precursors synthesized via the secretory pathway in transgenic plants.

14. Use of a propeptide according to claim 13 wherein the protein is a precursor of a plant defensin, or a hevein-type antimicrobial protein or is isolatable from the genus Impatiens.

15. Use of a propeptide as a cleavable linker in polyprotein precursors synthesized via the secretory pathway in transgenic plants wherein said propeptide linker is as defined in claim 4 or claim 5.

16. Use of a propeptide sequence rich in the small amino acids A, V, S and T and containing dipeptidic sequences consisting of either two acidic residues, two basic residues or one acidic and one basic residue as a cleavable linker sequence wherein said sequence is isolatable from a plant defensin or a hevein-type antimicrobial peptide.

17. A DNA construct comprising a DNA sequence comprising a promoter region operably linked to a plant derived signal sequence said signal sequence being operably linked to two or more protein encoding regions and a 3' terminator-region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide said propeptide providing a post-translational cleavage site.

18. A DNA construct according to claim 17 wherein said DNA sequence encoding said linker propeptide encodes an internal propeptide from the Ib-AMP gene, or the C-terminal propeptide from the Dm-AMP or from the Ac-AMP gene.

19. A DNA construct comprising a DNA sequence comprising a promoter region operably linked to two or more protein encoding regions and a 3' terminator-region wherein said protein encoding regions are separated from each other by a DNA sequence coding for a linker propeptide encoding the C-terminal propeptide from the Dm-AMP gene or the from the Ac-AMP gene said propeptide providing a post-translational cleavage site

20. A DNA construct according to claim 17, claim 18 or claim 19 wherein the DNA sequence encoding the linker propeptide additionally comprises one or more protease recognition sites at either or both ends thereof.

21. A vector comprising a DNA construct according to any of claims 17 to 20.

22. A transgenic plant transformed with a DNA construct or a vector according to claim 17 to 21.

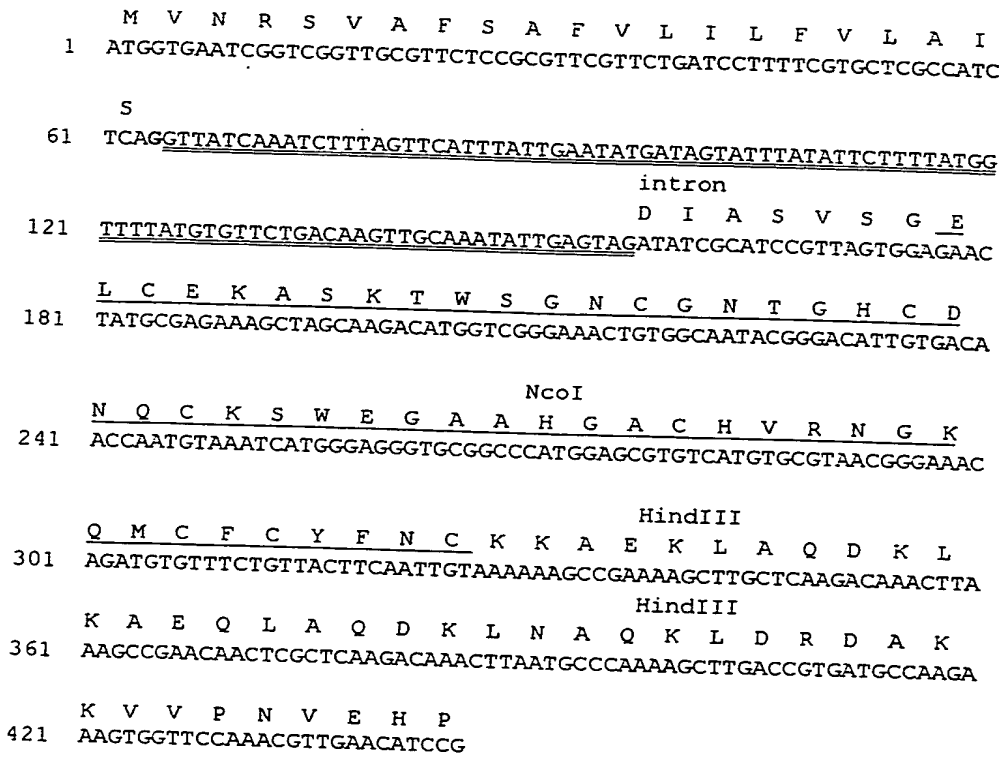

M V N R S V A F S A F V L I L F V L A I
1 ATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTTTTCGTGCTCGCCATC
S
61 TCAGGTTATCAAATCTTTAGTTCATTTATTGAATATGATAGTATTTATATTCTTTTATGG
intron
D I A S V S G E
121 TTTTATGTGTTCTGACAAGTTGCAAATATTGAGTAGATATCGCATCCGTTAGTGGAGAAC
L C E K A S K T W S G N C G N T G H C D
181 TATGCGAGAAAGCTAGCAAGACATGGTCGGGAAACTGTGGCAATACGGGACATTGTGACA
NcoI
N Q C K S W E G A A H G A C H V R N G K
241 ACCAATGTAAATCATGGGAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAAC
HindIII
Q M C F C Y F N C K K A E K L A Q D K L
301 AGATGTGTTTCTGTTACTTCAATTGTAAAAAAGCCGAAAAGCTTGCTCAAGACAAACTTA
HindIII
K A E Q L A Q D K L N A Q K L D R D A K
361 AAGCCGAACAACTCGCTCAAGACAAACTTAATGCCCAAAAGCTTGACCGTGATGCCAAGA
K V V P N V E H P
421 AAGTGGTTCCAAACGTTGAACATCCG

Figure 1

Construct #

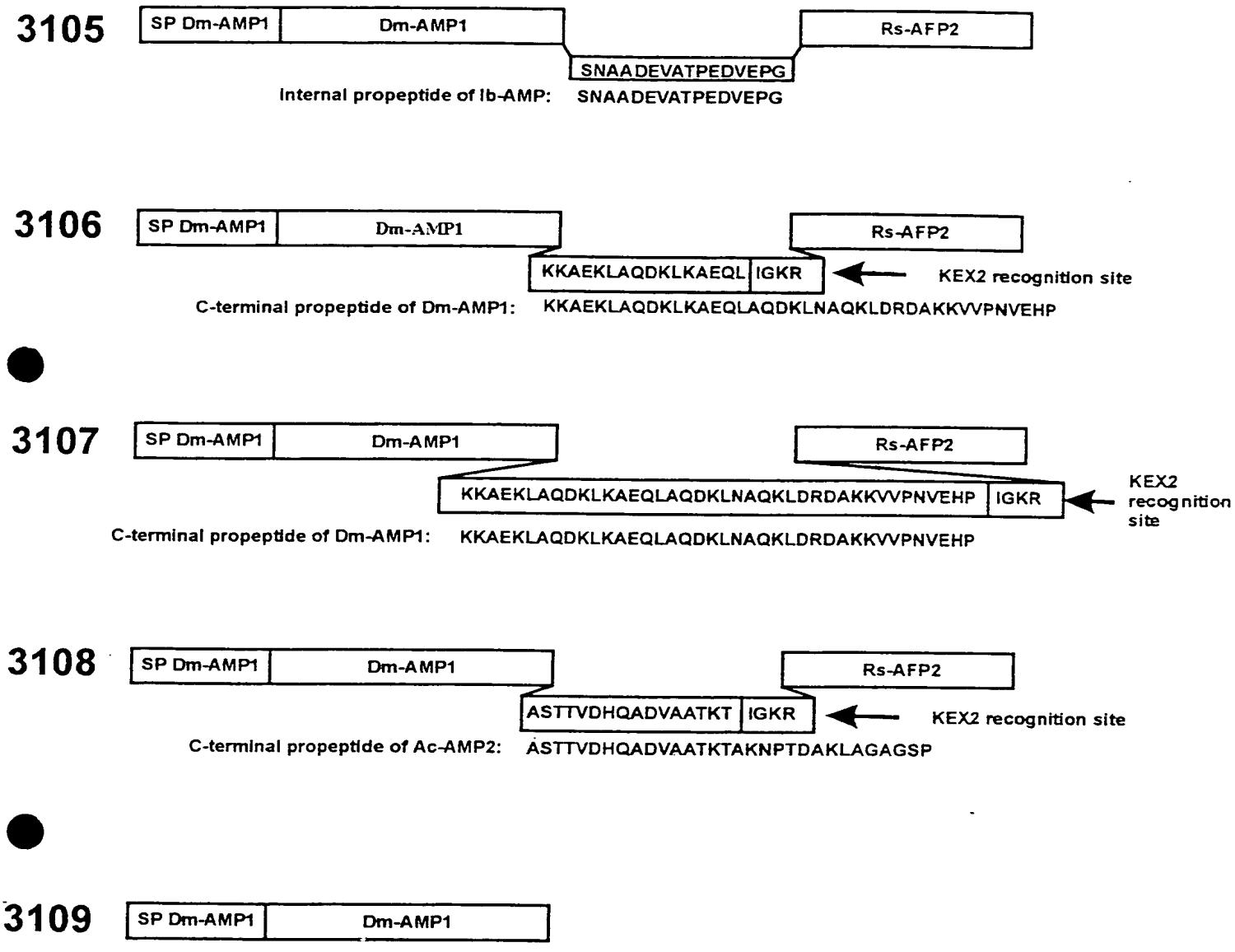
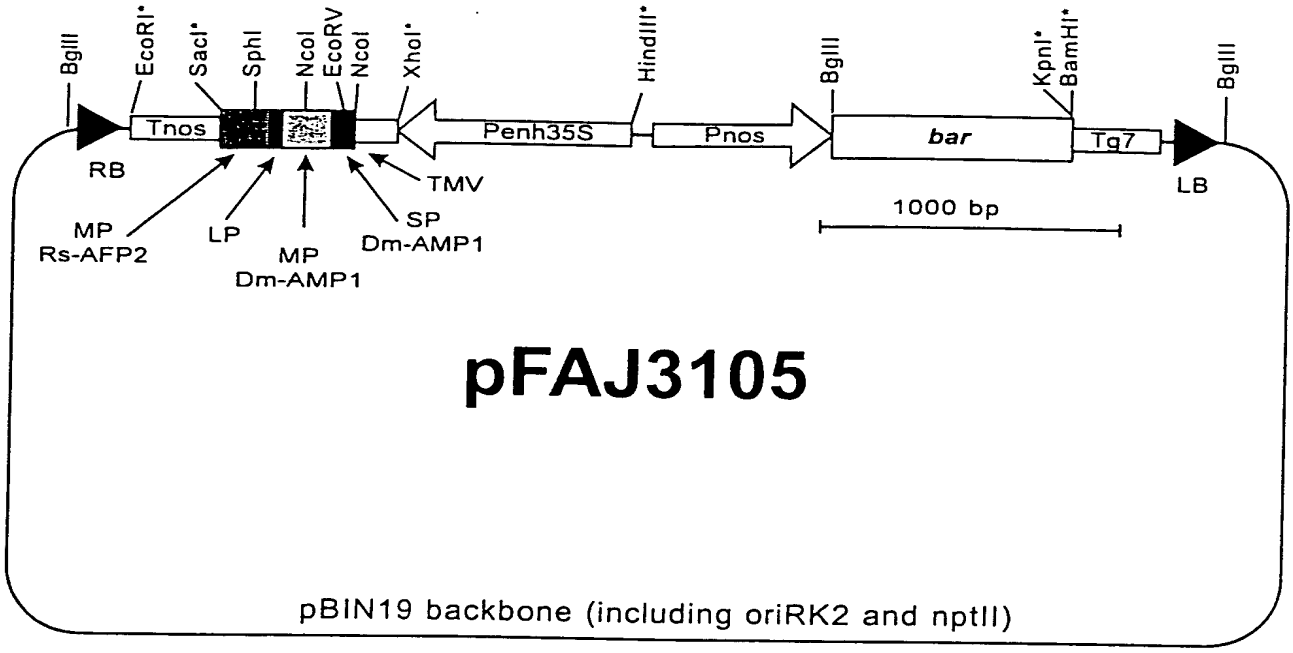



Figure 2

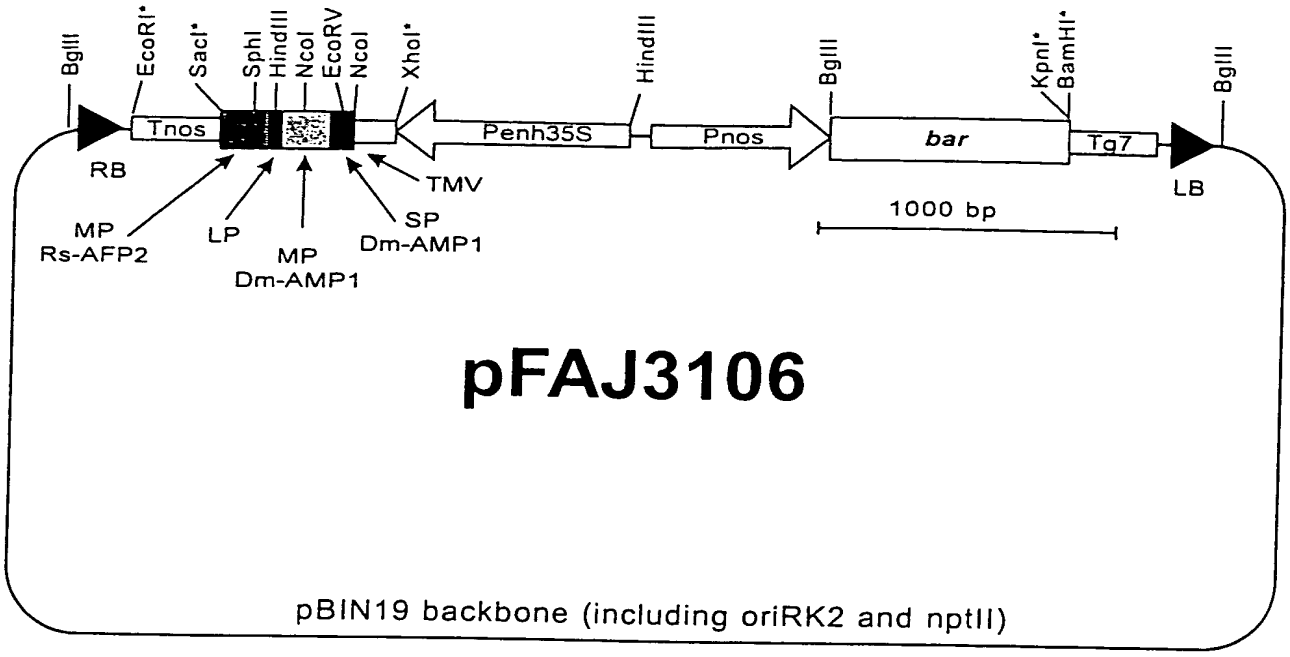



## Symbols

RB: right border of T-DNA
Tnos: terminator of T-DNA nopaline synthase gene
MP Rs-AFP2: mature protein domain of Rs-AFP2
LP: Ib-AMP internal propeptide
MP Dm-AMP1: mature protein domain of Dm-AMP1 cDNA
SP Dm-AMP1: signal peptide domain of Dm-AMP1 cDNA
TMV: tobacco mosaic virus 5' leader sequence
Penh35S: promotor of 35S RNA of cauliflower mosaic virus with duplicated enhancer region
Pnos: promotor of T-DNA nopaline synthase gene
bar: basta resistance encoding gene
Tg7: terminator of T-DNA gene 7
LB: left border of T-DNA

*: unique restriction site

Figure 3

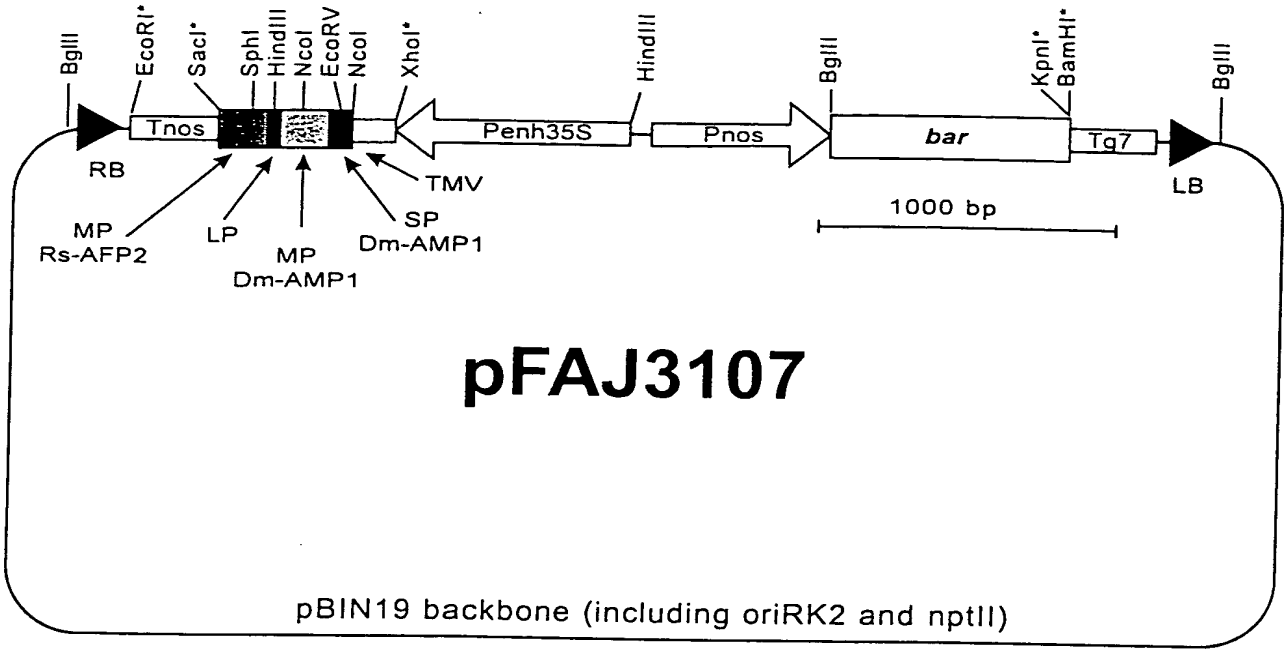



## Symbols

RB: right border of T-DNA
Tnos: terminator of T-DNA nopaline synthase gene
MP Rs-AFP2: mature protein domain of Rs-AFP2
LP: first 16 AA of Dm-AMP1 C-terminal propeptide and subtilisin-like protease recognition site IGKR
MP Dm-AMP1: mature protein domain of Dm-AMP1 cDNA
SP Dm-AMP1: signal peptide domain of Dm-AMP1 cDNA
TMV: tobacco mosaic virus 5' leader sequence
Penh35S: promotor of 35S RNA of cauliflower mosaic virus with duplicated enhancer region
Pnos: promotor of T-DNA nopaline synthase gene
bar: basta resistance encoding gene
Tg7: terminator of T-DNA gene 7
LB: left border of T-DNA

*: unique restriction site

Figure 4

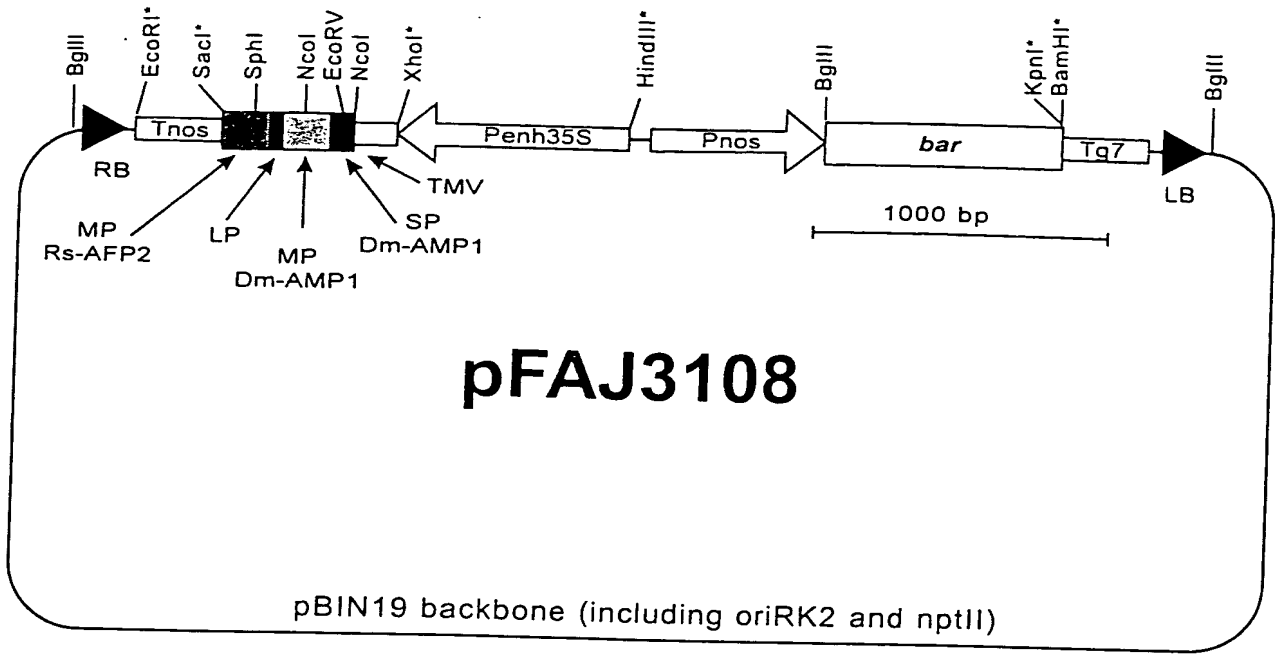



## Symbols

RB: right border of T-DNA
Tnos: terminator of T-DNA nopaline synthase gene
MP Rs-AFP2: mature protein domain of Rs-AFP2
LP: Dm-AMP1 C-terminal propeptidedomain and subtilisin-like protease recognition site IGKR
MP Dm-AMP1: mature protein domain of Dm-AMP1 cDNA
SP Dm-AMP1: signal peptide domain of Dm-AMP1 cDNA
TMV: tobacco mosaic virus 5' leader sequence
Penh35S: promotor of 35S RNA of cauliflower mosaic virus with duplicated enhancer region
Pnos: promotor of T-DNA nopaline synthase gene
bar: basta resistance encoding gene
Tg7: terminator of T-DNA gene 7
LB: left border of T-DNA

*: unique restriction site

Figure 5

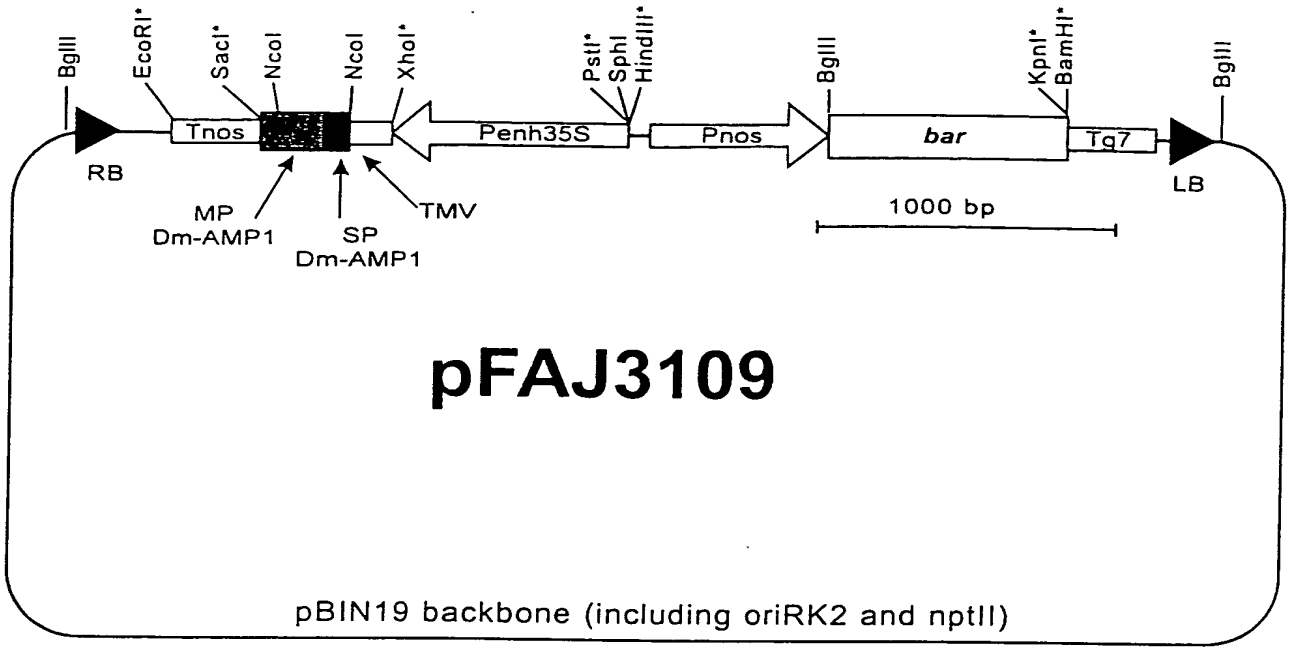



## Symbols

RB: right border of T-DNA
Tnos: terminator of T-DNA nopaline synthase gene
MP Rs-AFP2: mature protein domain of Rs-AFP2
LP: first 16 AA of Ac-AMP2 C-terminal propeptide domain and subtilisin-like protease recognition site IGKR
MP Dm-AMP1: mature protein domain of Dm-AMP1 cDNA
SP Dm-AMP1: signal peptide domain of Dm-AMP1 cDNA
TMV: tobacco mosaic virus 5' leader sequence
Penh35S: promotor of 35S RNA of cauliflower mosaic virus with duplicated enhancer region
Pnos: promotor of T-DNA nopaline synthase gene
bar: basta resistance encoding gene
Tg7: terminator of T-DNA gene 7
LB: left border of T-DNA

*: unique restriction site

Figure 6




## Symbols

RB: right border of T-DNA
Tnos: terminator of T-DNA nopaline synthase gene
MP Dm-AMP1: mature protein domain of Dm-AMP1
SP Dm-AMP1: signal peptide domain of Dm-AMP1 cDNA
TMV: tobacco mosaic virus 5' leader sequence
Penh35S: promotor of 35S RNA of cauliflower mosaic virus with duplicated enhancer region
Pnos: promotor of T-DNA nopaline synthase gene
bar: basta resistance encoding gene
Tg7: terminator of T-DNA gene 7
LB: left border of T-DNA

*: unique restriction site

Figure 7

pFAJ3105

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

TTCAATTGTTCCAACGCTGCTGACGAGGTGGCTACCCCAGAGGACGTGGAGCCAGGACAG

F N C **S N A A D E V A T P E D V E P G** Q

AAGTTGTGCCAAAGGCCAAGTGGGACATGGTCAGGAGTCTGTGGAAACAATAACGCATGC

K L C Q R P S G T W S G V C G N N N A C

AAGAATCAGTGCATTAGACTTGAGAAAGCACGACATGGATCTTGCAACTATGTCTTCCCA

K N Q C I R L E K A R H G S C N Y V F P

SacI

GCTCACAAGTGTATCTGCTACTTTCCTTGTTAATAGGAGCTC

A H K C I C Y F P C - -

Figure 8

pFAJ3106

XhoI

```
CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT
              NcoI
ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT
                  M  V  N  R  S  V  A  F  S  A  F  V  L  I  L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC
 F  V  L  A  I  S  D  I  A  S  V  S  G  E  L  C  E  K  A  S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG
 K  T  W  S  G  N  C  G  N  T  G  H  C  D  N  Q  C  K  S  W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAACACATGTGTTTCTGTTAC
 E  G  A  A  H  G  A  C  H  V  R  N  G  K  H  M  C  F  C  Y

TTCAATTGTAAAAAAGCCGAAAAGCTTGCTCAAGACAAACTTAAAGCCGAACAACTCATC
 F  N  C  K  K  A  E  K  L  A  Q  D  K  L  K  A  E  Q  L  I

GGAAAGAGGCAGAAGTTGTGCCAAAGGCCAAGTGGGACATGGTCAGGAGTCTGTGGAAAC
 G  K  R  Q  K  L  C  Q  R  P  S  G  T  W  S  G  V  C  G  N

AATAACGCATGCAAGAATCAGTGCATTAGACTTGAGAAAGCACGACATGGATCTTGCAAC
 N  N  A  C  K  N  Q  C  I  R  L  E  K  A  R  H  G  S  C  N
                                                  SacI
TATGTCTTCCCAGCTCACAAGTGTATCTGCTACTTTCCTTGTTAATAGGAGCTC
 Y  V  F  P  A  H  K  C  I  C  Y  F  P  C  -  -
```

Figure 9

pFAJ3107

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

TTCAATTGTAAAAAAGCCGAAAAGCTTGCTCAAGACAAACTTAAAGCCGAACAACTCGCT

F N C K K A E K L A Q D K L K A E Q L A

CAAGACAAACTTAATGCCCAAAAGCTTGACCGTGATGCCAAGAAAGTGGTTCCAAACGTT

Q D K L N A Q K L D R D A K K V V P N V

GAACATCCGATCGGAAAGAGGCAGAAGTTGTGCCAAAGGCCAAGTGGGACATGGTCAGGA

E H P I G K R Q K L C Q R P S G T W S G

GTCTGTGGAAACAATAACGCATGCAAGAATCAGTGCATTAGACTTGAGAAAGCACGACAT

V C G N N N A C K N Q C I R L E K A R H

GGATCTTGCAACTATGTCTTCCCAGCTCACAAGTGTATCTGCTACTTTCCTTGTTAATAG

G S C N Y V F P A H K C I C Y F P C - -

SacI

GAGCTC

Figure 10

pFAJ3108

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAACGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

TTCAATTGTGCCAGTACTACTGTGGATCACCAAGCTGATGTTGCTGCCACCAAAACTATC

F N C A S T T V D H Q A D V A A T K T I

GGAAAGAGGCAGAAGTTGTGCCAAAGGCCAAGTGGGACATGGTCAGGAGTCTGTGGAAAC

G K R Q K L C Q R P S G T W S G V C G N

AATAACGCATGCAAGAATCAGTGCATTAGACTTGAGAAAGCACGACATGGATCTTGCAAC

N N A C K N Q C I R L E K A R H G S C N

SacI

TATGTCTTCCCAGCTCACAAGTGTATCTGCTACTTTCCTTGTTAATAGGAGCTC

Y V F P A H K C I C Y F P C - -

Figure 11

pFAJ3109

XhoI

CTCGAGTATTTTTACAACAATTACCAACAACAACAAACAACAAACAACATTACAATTACT

NcoI

ATTTACAATTACACCATGGTGAATCGGTCGGTTGCGTTCTCCGCGTTCGTTCTGATCCTT

M V N R S V A F S A F V L I L

TTCGTGCTCGCCATCTCAGATATCGCATCCGTTAGTGGAGAACTATGCGAGAAAGCTAGC

F V L A I S D I A S V S G E L C E K A S

AAGACGTGGTCGGGCAACTGTGGCAACACGGGACATTGTGACAACCAATGTAAATCATGG

K T W S G N C G N T G H C D N Q C K S W

GAGGGTGCGGCCCATGGAGCGTGTCATGTGCGTAATGGGAAACACATGTGTTTCTGTTAC

E G A A H G A C H V R N G K H M C F C Y

SacI

TTCAATTGTTGAGCTC

F N C -

Figure 12

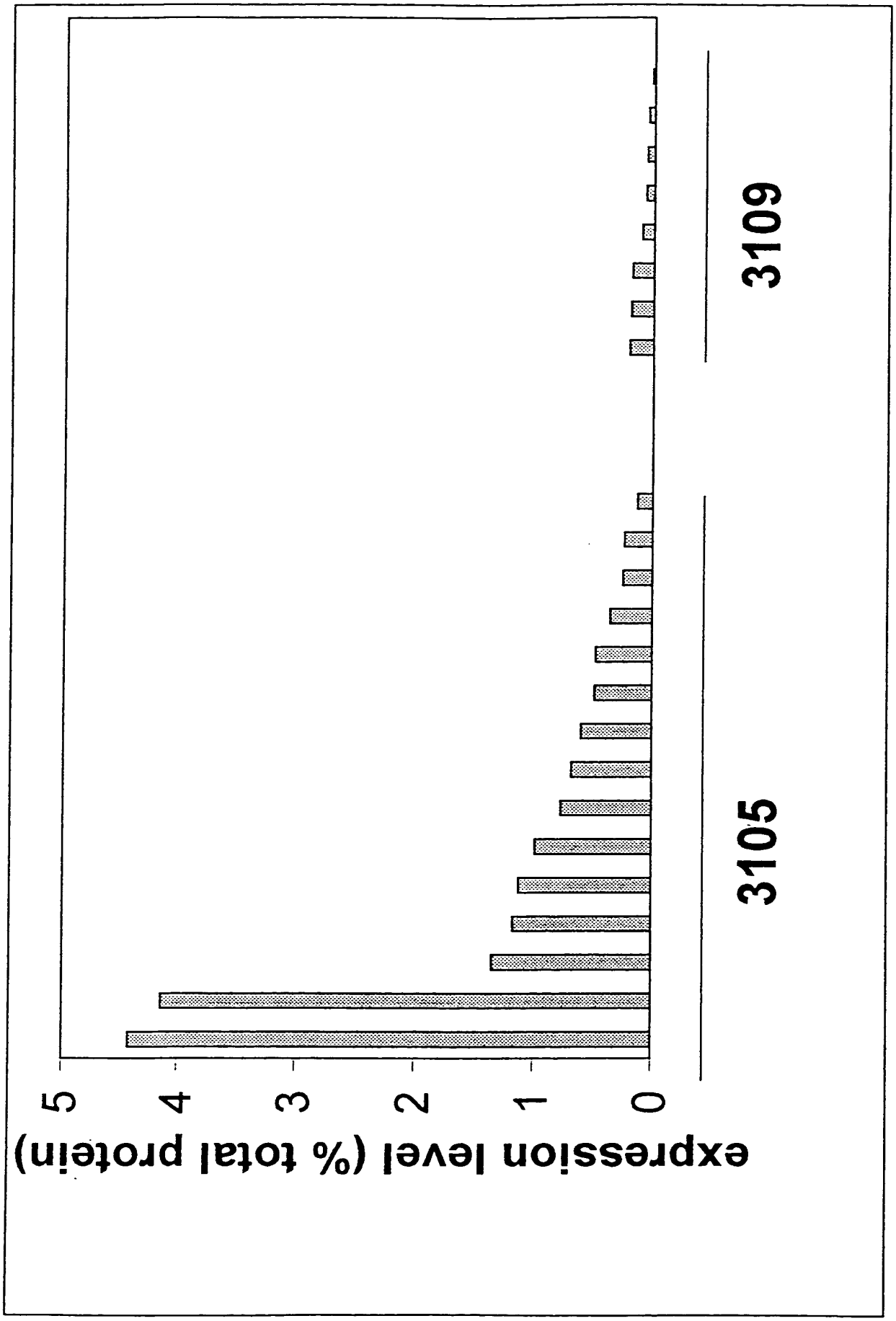
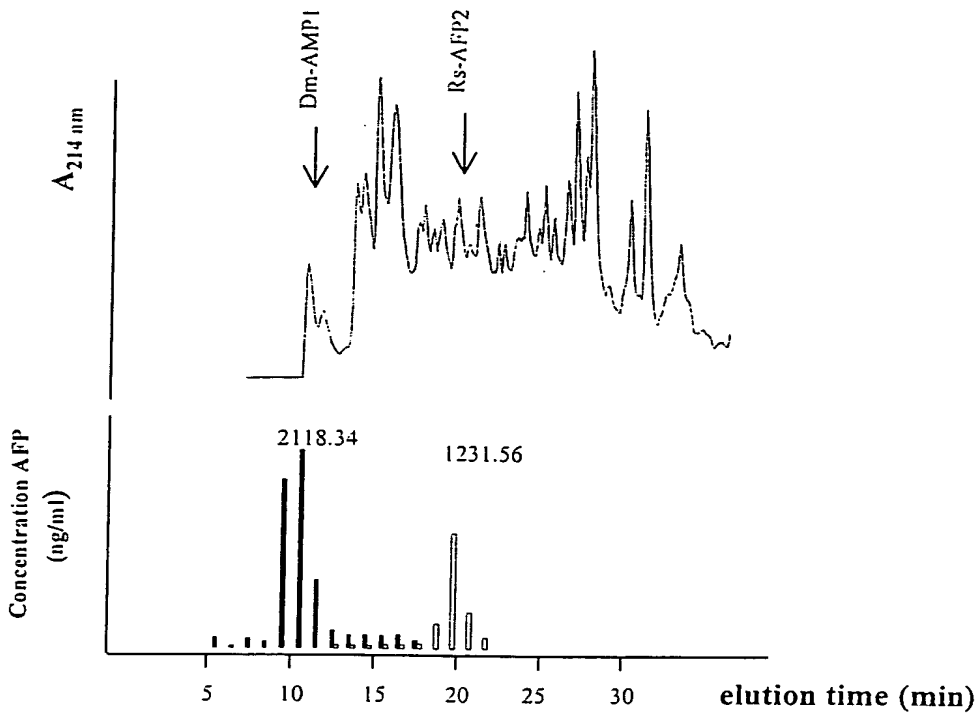



Figure 13

A


Dm-AMP1
Rs-AFP2
$A_{214\ nm}$
2118.34
1231.56
Concentration AFP (ng/ml)
5
10
15
20
25
30
elution time (min)

B

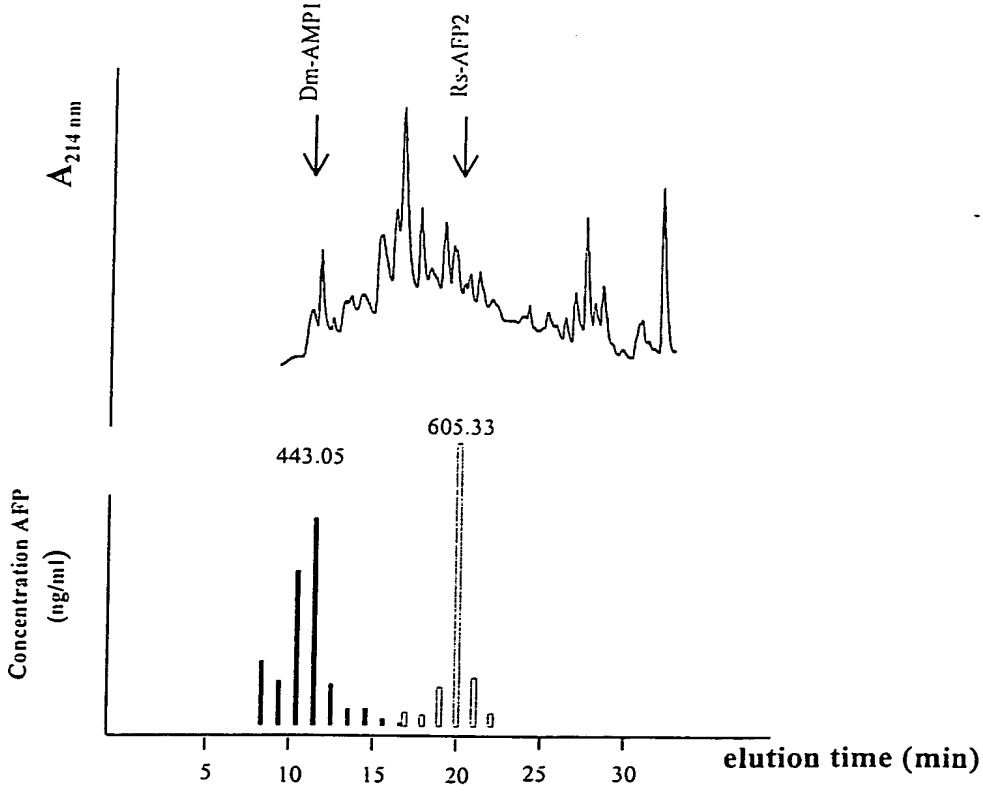

Dm-AMP1
Rs-AFP2
$A_{214\ nm}$
605.33
443.05
Concentration AFP (ng/ml)
5
10
15
20
25
30
elution time (min)

Figure 14

PCT NO : 99/02716

FORM 23/77 : 17.8.99

AGENT : Zeneca Agrochemicals